



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### From state distinguishability to effective bulk locality

**Citation for published version:**

Simon, J & Lashkari, N 2014, 'From state distinguishability to effective bulk locality' Journal of High Energy Physics, vol. 14, no. 04. DOI: 10.1007/JHEP06(2014)038

**Digital Object Identifier (DOI):**

[10.1007/JHEP06\(2014\)038](https://doi.org/10.1007/JHEP06(2014)038)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Journal of High Energy Physics

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# From state distinguishability to effective bulk locality

Nima Lashkari<sup>a,b</sup> and Joan Simón<sup>c</sup>

<sup>a</sup>*Stanford Institute for Theoretical Physics and Department of Physics, Stanford University, Stanford, CA 94305-4060, U.S.A.*

<sup>b</sup>*Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, B.C., V6T 1W9, Canada*

<sup>c</sup>*School of Mathematics and Maxwell Institute for Mathematical Sciences, King's Buildings, Edinburgh EH9 3JZ, United Kingdom*

E-mail: [nimal@stanford.edu](mailto:nimal@stanford.edu), [j.simon@ed.ac.uk](mailto:j.simon@ed.ac.uk)

**ABSTRACT:** We provide quantitative evidence that the emergence of an effective notion of spacetime locality in black hole physics is due to restricting to the subset of observables that are unable to resolve black hole microstates from the maximally entangled state. We identify the subset of observables in the full quantum theory that can distinguish microstates, and argue that any measurement of such observables involves either long times or large energies, both signaling the breaking down of effective field theory where locality is manifest. We discuss some of the implications of our results for black hole complementarity and the existence of black hole interiors.

**KEYWORDS:** Gauge-gravity correspondence, AdS-CFT Correspondence, Black Holes

**ARXIV EPRINT:** [1402.4829](https://arxiv.org/abs/1402.4829)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Black holes vs quantum statistical mechanics</b>	<b>2</b>
<b>3</b>	<b>Distinguishing quantum states</b>	<b>4</b>
3.1	Expectation value of operators	4
3.2	The use of typicality on expectation values	6
3.3	Measure of distinguishability	8
3.4	Coarse-grained observables	11
<b>4</b>	<b>Lessons for quantum gravity</b>	<b>12</b>
4.1	Breaking down of effective field theory	13
4.2	Effective emergence of locality	14
<b>A</b>	<b>Ensembles, averages &amp; variances</b>	<b>18</b>
<b>B</b>	<b>Proof of theorem 3.2</b>	<b>20</b>
<b>C</b>	<b>Proof of theorem 3.3</b>	<b>22</b>

---

## 1 Introduction

One of the main open questions in black hole physics is the compatibility of unitarity in the entire quantum theory with the emergence of some notion of spacetime locality, the latter being manifest in any effective field theory (EFT) description.

In this paper we explore the emergence of some notion of *effective bulk locality* in light of the connection between the resolution of the information paradox, and the modern formulation of thermalization in terms of *entanglement* and *typicality*. In a nutshell, just as the emergence of *effective thermalization* in quantum statistical mechanics requires some notion of coarse-graining, we propose that the emergence of effective bulk locality is due to the restriction to a subset of observables which cannot resolve black hole microstates from their maximally entangled averages.

Using tools from quantum information theory, we prove that a given observable  $\mathcal{O}$ , or a collection of them, cannot distinguish a random pure microstate in a microcanonical ensemble  $\mathcal{H}_E$  of dimension  $d_E$  from the maximally entangled state  $\Omega_E = \frac{\mathbb{I}_E}{d_E}$  unless the number of different outcomes of the operator  $N(\mathcal{O})$  scales as  $\sqrt{d_E}$ . Furthermore, whenever  $N(\mathcal{O}) \sim \sqrt{d_E}$ , we prove, in a simple quantum mechanical model, that any quantum measurement would require either a very long time or involve a very large energy to achieve the accuracy required to distinguish these states. Either way, this points to a breakdown

in the EFT description. Alternatively, we show that any measurement involving a finite amount of resources is necessarily coarse-grained.

In an AdS/CFT set-up, our results provide mathematical evidence that the operators belonging to the bulk low energy effective theory are coarse-grained, in the sense that they are unable to distinguish random pure states from the maximally mixed one unless you wait for exponentially long time. Furthermore, they also support the idea that only low point correlators of such operators admit a *local* bulk semi-classical interpretation compatible with the EFT description. Taking this perspective, we further comment on some implications for quantum gravity: mainly the consistency with black hole complementarity and the relation with recent discussions regarding the (non-)existence of a classical black hole interior.

## 2 Black holes vs quantum statistical mechanics

Hawking established that black holes radiate with a thermal spectrum [1] by performing a 2-point function calculation in a framework that we shall henceforth refer to as effective field theory (EFT): he considered a bulk quantum field  $\phi(x)$  propagating in a non-dynamical black hole background

$$\langle \psi_{\text{BH}} | \phi(x_1) \phi(x_2) | \psi_{\text{BH}} \rangle_{\text{EFT}}, \quad (2.1)$$

where the subscript EFT refers to the replacement of the black hole quantum state  $|\psi_{\text{BH}}\rangle$  by the classical black hole background.

This description is manifestly local, a property that is expected to hold due to the equivalence principle, but is hard to derive in any quantum theory of gravity. Locality appears to be in contradiction with the unitarity of quantum mechanics, and this is commonly referred to as the *information paradox*. One aspect of this paradox is directly related to the *problem of thermalization* in quantum mechanics: given a *closed* quantum system in an initial pure state, unitary evolution never gives rise to a mixed state such as the Gibbs state. The latter is a fundamental problem for the foundations of quantum statistical mechanics (QSM).

How do we understand thermalization then? In QSM, thermalization is always viewed as a consequence of *coarse-graining* the description of our physical system. In modern language, one expects *entanglement* and *typicality* to explain why QSM is an accurate description of nature at a macroscopic level. For example, consider a composite Hilbert space  $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_B$ , where  $\mathcal{H}_S$  is the subspace that we will measure and  $\mathcal{H}_B$  stands for the bath. Given an initial quantum state  $\rho \in \mathcal{H}$ , its reduced density matrix  $\rho_S$

$$\rho_S = \text{tr}_{\mathcal{H}_B} \rho \quad (2.2)$$

is the only information that we have access to. Quantum entanglement is responsible for encoding the information about correlations between the subsystems whereas typicality is responsible for the apparent universality in our measurements. More precisely, if the dimensions of these Hilbert spaces satisfy  $d_S \ll \sqrt{d_B}$ , in the thermodynamic limit  $d_B \rightarrow \infty$ ,

the deviation of  $\rho_S$  from the maximally entangled state  $\Omega_S = \frac{\mathbb{I}_S}{d_S} \in \mathcal{H}_S$  is suppressed [2, 3]. That is

$$\|\rho_S - \Omega_S\| \rightarrow 0 \quad \text{when} \quad \frac{d_S}{\sqrt{d_B}} \rightarrow 0 \quad (2.3)$$

where  $\|\dots\|$  stands for trace distance.<sup>1</sup>

In this work we take the perspective that such coarse-graining in holographic field theories is responsible for the emergence of a notion of *effective locality* (and consequently, the causal structure of spacetime), which is an assumption in Hawking's calculation. To advance our main philosophy, bulk locality emerges only for the subset of observables that fail to distinguish a thermal state, or  $\Omega_S$  in a microcanonical set-up, from the actual microstate of the black hole. In gravitational physics having holographic duals, the coarse-graining appears as a result of restricting to the subset of low energy observables  $\mathcal{A}_{\text{low}}$  belonging to the effective field theory. These correspond to the light sector of the holographic theory in the terminology introduced in [4]. We will discuss this point more thoroughly in section 4.

With this motivation in mind, the main task we undertake is to identify the subset of observables  $\mathcal{O}$  that can distinguish pure microstates  $\rho_i = |\psi_i\rangle\langle\psi_i|$  from the thermal density matrix  $\rho_{\text{BH}}$  (the semi-classical description of the black hole)

$$\text{tr}(\rho_i \mathcal{O}) \neq \text{tr}(\rho_{\text{BH}} \mathcal{O}), \quad (2.4)$$

in a given large but *finite* dimensional Hilbert space  $\mathcal{H}$ .<sup>2</sup> In an exact quantum theory of gravity, such as in an AdS/CFT context, we quantify the deviations of correlators computed in black hole microstates  $|\psi_i\rangle$  from the effective field theory answer. These deviations are expected to be of the form [5, 6]

$$\langle\psi_i|\mathcal{O}_1(x_1)\dots\mathcal{O}_N(x_N)|\psi_i\rangle_{\text{exact}} = \langle\psi_{\text{BH}}|\mathcal{O}_1(x_1)\dots\mathcal{O}_N(x_N)|\psi_{\text{BH}}\rangle_{\text{exact}} + \mathcal{O}(e^{-(S_{\text{BH}}-N)}) \quad (2.5)$$

where  $e^{S_{\text{BH}}} = d_E$  is the dimension of the microcanonical Hilbert space  $\mathcal{H}_E$  of black hole microstates. The  $e^{-S_{\text{BH}}}$  exponential suppression is expected from semiclassical gravity [7] and statistical considerations [8]. The last term in the right hand side is currently a vague way of parameterising the fact that depending on the nature of the correlator, i.e. the number  $N$  of *insertions* and the properties of the individual *probe* operators  $\mathcal{O}_i$ , such corrections may not be subleading [6, 8, 9]. Our main goal is to make this statement precise in quantum mechanics.

Even though our analysis is in the full quantum theory, it concerns the robustness of Hawking's EFT framework because whenever thermal answers are not sharply peaked due to large variances, the notion of bulk locality is not expected to hold. We stress this breaking down in the EFT description is, a priori, on top of the well established fact that low point correlators, such as the 2-pt function in Hawking's original calculation, must break down at large times, i.e. for times of the order of the black hole evaporation time scale or the Poincaré recurrence time. The latter can be made particularly precise in an AdS/CFT context [7].

---

<sup>1</sup> $\|A\| = \text{tr}(|A|) = \sum_i |a_i|$  where  $a_i$ 's are eigenvalues of  $A$ .

<sup>2</sup>Strictly speaking, our proofs involve the maximally entangled state rather than the Gibbs state. Extending our tools for the canonical ensemble is a hard open problem we will not attempt here.

### 3 Distinguishing quantum states

Consider a finite dimensional subspace  $\mathcal{H}_E \subset \mathcal{H}$  of dimension  $d_E$  consisting of all pure states  $\psi = |\psi\rangle\langle\psi|$  that live in the microcanonical ensemble of energy  $[E - \delta E, E + \delta E]$ . We will assume the Hamiltonian describing the unitary time evolution of the system has *non-degenerate energy gaps*. This is a condition on the spectrum stating that the equality  $E_k - E_l = E_m - E_n$  can only be satisfied either by  $E_k = E_l$  and  $E_m = E_n$ , or by  $E_k = E_m$  and  $E_l = E_n$ .<sup>3</sup>

In this section, we identify the necessary condition for a set of observables  $\mathcal{A}$  to distinguish a random pure state  $\psi \in \mathcal{H}_E$  from the maximally mixed state in  $\mathcal{H}_E$ . We argue that an actual quantum mechanical measurement of any such observable requires large resources, i.e. very long times or very high energies. Alternatively, any finite time and energy resourced quantum measurement is equivalent to a *coarse-grained* observable for which random pure states still appear entangled.

#### 3.1 Expectation value of operators

One possibility to quantify the difference between quantum states  $\psi \in \mathcal{H}_E$  is to measure the expectation value of some operator  $A$ . We can study this either at fixed time, by averaging over the entire set of states, or by averaging over time.

To study these questions, it is instructive to think of  $\langle\psi|A|\psi\rangle$  as a random variable  $X$  with uniform distribution on  $\mathcal{H}_E$  or over the positive real line  $\mathbb{R}^+$ , respectively [11].<sup>4</sup> Applying Chebyshev's inequality

$$\text{Prob}(|X - \langle X \rangle| \geq \delta x) \leq \frac{\sigma_X^2}{\delta x^2} \quad (3.1)$$

provides an upper bound on the probability for  $X$  to differ from its expectation value  $\langle X \rangle$  by some quantity  $\delta x$ .<sup>5</sup>

**Ensemble average.** Consider a random state  $\psi \in \mathcal{H}_E$ . Since the random variable  $X_\psi = \langle\psi|A|\psi\rangle$  is uniformly distributed over  $\mathcal{H}_E$ , its average value  $\langle A \rangle_E$  will equal

$$\langle A \rangle_E = \frac{1}{d_E} \text{tr} A = \text{tr}(A \Omega_E) \quad \text{with} \quad \Omega_E = \frac{\mathbb{I}_E}{d_E}. \quad (3.2)$$

Thus, the ensemble average always equals the expectation value of the operator  $A$  in the maximally entangled state  $\Omega_E$  in  $\mathcal{H}_E$ .

The variance of the random variable [8, 12]

$$\sigma_A^2 \equiv \langle (\langle\psi|A|\psi\rangle - \langle A \rangle_\psi)^2 \rangle_\psi = \frac{1}{d_E + 1} \left( \frac{1}{d_E} \text{tr}(A^2) - \left( \frac{\text{tr}(A)}{d_E} \right)^2 \right) \quad (3.3)$$

<sup>3</sup>This assumption is believed to be correct for almost all interacting Hamiltonians. The addition of an arbitrarily small random term to the Hamiltonians lifts all degeneracies. All our results continue to hold as long as no energy gap is highly degenerate [10].

<sup>4</sup>For a proper definition of the different ensemble averages discussed below, see appendix A.

<sup>5</sup>The bound (3.1) is only meaningful whenever  $\frac{\sigma_X^2}{\delta x^2} \leq 1$ .

is inversely proportional to the ensemble dimension, i.e. exponentially suppressed in the entropy of the system since,  $(d_E + 1)^{-1} \approx e^{-S_E}$ .

Chebyshev's inequality (3.1) provides an upper bound on the probability of observing a large deviation from the averaged expectation values

$$\text{Prob}(|\langle \psi | A | \psi \rangle - \text{tr}(A \Omega_E)| \geq \delta a) \leq \frac{\sigma_A^2}{(\delta a)^2}. \quad (3.4)$$

**Time average.** Given the existence of recurrences in finite dimensional Hilbert spaces, time evolution of any typical state  $\psi$  becomes arbitrarily close to any other state. It is important to determine the probability that a given state differs from its equilibrium configuration at a given instant of time. To study this, we choose a different ensemble and define a random variable  $X_t^\psi = \langle \psi(t) | A | \psi(t) \rangle$  with uniform distribution over  $t \in [0, \infty)$ . Given an initial random state expanded in energy eigenbasis

$$|\psi(0)\rangle = \sum_E c_E |E\rangle \in \mathcal{H}_E, \quad \text{with} \quad \sum_E |c_E|^2 = 1, \quad (3.5)$$

it evolves to  $|\psi(t)\rangle = \sum_E e^{-iEt} c_E |E\rangle$ . Its time-average equals

$$\omega_\psi \equiv \langle \psi(t) \rangle_t = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt |\psi(t)\rangle \langle \psi(t)| = \sum_E |c_E|^2 |E\rangle \langle E|. \quad (3.6)$$

Notice that in this ensemble the *equilibrium* configuration  $\omega_\psi$  depends explicitly on the initial state  $|\psi(0)\rangle$  and is obtained by setting to zero all off-diagonal elements in  $|\psi(t)\rangle \langle \psi(t)|$  in the energy eigenbasis. The latter is due to the dephasing occurring when averaging over time.

The variance  $\sigma_A^2(\psi)$  in this random variable is

$$\begin{aligned} \sigma_A^2(\psi) &= \left\langle \left( \text{tr}(A \psi(t)) - \text{tr}(A \omega_\psi) \right)^2 \right\rangle_t \\ &= \sum_{E, E'} |c_E|^2 |c_{E'}|^2 |\langle E' | A | E \rangle|^2 - \sum_E |c_E|^4 |\langle E | A | E \rangle|^2, \end{aligned} \quad (3.7)$$

where we assumed the hamiltonian has a non-degenerate energy gap spectrum. Chebyshev's inequality implies

$$\text{Prob}(|\langle \psi(t) | A | \psi(t) \rangle - \text{tr}(A \omega_\psi)| \geq \delta a) \leq \frac{\sigma_A^2(\psi)}{(\delta a)^2}. \quad (3.8)$$

**Ensemble vs time averages.** The equilibrium configuration  $\omega_\psi$  depends on the initial state and one can equally ask how large the probability is for a random equilibrium state to differ from the ensemble average (3.2). This can be efficiently computed from previous calculations simply by replacing the arbitrary operator  $A$  in the ensemble average discussion by  $\sum_E |E\rangle \langle E| A |E\rangle \langle E|$ . When we do this, the term  $\text{tr}(A \Omega_E)$  remains invariant, whereas the random expectation value  $\langle \psi | A | \psi \rangle$  becomes

$$\langle \psi | \left( \sum_E |E\rangle \langle E| A |E\rangle \langle E| \right) | \psi \rangle = \sum_E |\langle E | \psi \rangle|^2 \langle E | A | E \rangle = \text{tr}(A \omega_\psi). \quad (3.9)$$

Replacing this into (3.4), we obtain

$$\text{Prob}(|\text{tr}(A\omega_\psi) - \text{tr}(A\Omega_E)| \geq \delta a) \leq \frac{\tilde{\sigma}_A^2}{(\delta a)^2}, \quad (3.10)$$

with the new variance  $\tilde{\sigma}_A^2$  given by

$$\tilde{\sigma}_A^2 \equiv \langle (\text{tr}(A\omega_\psi) - \text{tr}(A\Omega_E))^2 \rangle_\psi. \quad (3.11)$$

### 3.2 The use of typicality on expectation values

Our goal is to show the above probabilities are suppressed in the dimensionality  $d_E$  of the Hilbert space  $\mathcal{H}_E$ . In the process, we learn which properties the operator  $A$  must satisfy to violate this conclusion.

To study these issues we will apply *typicality* arguments. These follow from the phenomenon denoted as concentration of measure in the mathematical literature [13, 14]. For our purposes, it is sufficient to point out that on a  $d$ -dimensional sphere  $\mathcal{S}^d$  of unit radius almost all points are within geodesic distance  $\frac{1}{\sqrt{d}}$  from its equator. More mathematically, given an spherical cap  $C(\epsilon)$  located a distance  $0 < \epsilon < 1$  from the center of the sphere, the isoperimetric inequality allows us to bound its normalised measure by

$$\sigma(C(\epsilon)) = \frac{\text{Area}(C(\epsilon))}{\text{Area}(\mathcal{S}^d)} \leq (\sqrt{1 - \epsilon^2})^{d+1} \leq e^{-(d+1)\epsilon^2/2}. \quad (3.12)$$

It follows from this inequality that the probability for a random point on this sphere to belong to a band  $B_\epsilon$  around the sphere excluding the pair of caps  $C(\epsilon)$  equals

$$\sigma(B_\epsilon) \geq 1 - 2e^{-(d+1)\epsilon^2/2}. \quad (3.13)$$

Our claim follows in the limit of large  $d$ .

These geometric facts become relevant for the foundations of quantum statistical mechanics when we consider subspaces  $\mathcal{H}_E$  of the Hilbert space with a large dimension. Indeed, from (3.5), a random pure state in  $\mathcal{H}_E$  corresponds to a random point on a sphere of dimension  $d = 2d_E - 1$  parameterised by the complex vector components  $c_E$ .

The application of these ideas to expectation values, i.e. functions over this sphere, is known as *Levy's lemma*.

**Lemma 3.1** *Levy's lemma: given a bounded function  $f(\psi)$  defined over the set of pure states  $\psi \in \mathcal{H}_E$ , for any such random state  $\psi$  and any  $\epsilon > 0$*

$$\text{Prob}(|f(\psi) - \langle f(\psi) \rangle_\psi| \geq \epsilon) \leq 2e^{-4c d_E \epsilon^2 / \lambda^2} \quad (3.14)$$

where  $c = (18\pi^3)^{-1}$ ,  $d_E$  is the dimension of  $\mathcal{H}_E$ , and  $\lambda = \sup |\nabla f|$  is the Lipschitz constant of the function  $f(\psi)$ .

See [13, 14] for a proof.



**Ensemble average.** As a first application of Levy's lemma, consider the random variable  $X_\psi$ . Its uniform probability distribution is precisely the normalised measure over the sphere describing the microstates in our microcanonical ensemble  $\mathcal{H}_E$ . Thus, expectation values of operators  $A$  are functions over this sphere, i.e.  $f(\psi) = \langle \psi | A | \psi \rangle$ , satisfying  $\langle f(\psi) \rangle_\psi = \text{tr}(A \Omega_E)$ . The Lipschitz constant  $\lambda$  controls how probable large deviations are. It was proved in [2, 3] that

$$\lambda = \sup |\nabla \langle \psi | A | \psi \rangle| \leq 2 \|A\|. \quad (3.15)$$

Thus, unless the largest eigenvalue of  $A$  scales like  $\sqrt{d_E}$ , we can conclude that such probability is exponentially suppressed in  $d_E$ .

**Time average.** The variance  $\sigma_A^2(\psi)$  in (3.16) is bounded above according to [11, 15]

$$\sigma_A^2(\psi) \leq \frac{\Delta_A^2}{4} \text{tr}(\omega_\psi^2). \quad (3.16)$$

This upper bound has two pieces. The quantity  $\text{tr}(\omega_\psi^2)$  is the *purity* of the equilibrium state, it equals  $\sum_E |c_E|^4$  and manifestly depends on the state. The quantity  $\Delta_A^2$  is the difference between the largest and the smallest eigenvalues of  $A$  restricted to  $\mathcal{H}_E$

$$\Delta_A = \max_{\psi \in \mathcal{H}_E} \langle \psi | A | \psi \rangle - \min_{\psi \in \mathcal{H}_E} \langle \psi | A | \psi \rangle. \quad (3.17)$$

It manifestly depends on the operator  $A$ .

A more sophisticated application of Levy's lemma (3.14) in [16] allows to prove

$$\text{Prob} \left( \text{tr}(\omega_\psi^2) > \frac{4}{d_E} \right) \leq 2e^{-\tilde{c}\sqrt{d_E}}, \quad (3.18)$$

where  $\tilde{c} = \frac{(\log 2)^2}{72\pi^3} \approx 10^{-4}$ . Thus,  $\text{tr}(\omega_\psi^2)$  is typically suppressed as  $\frac{1}{d_E}$ . Using this result in (3.8), we learn that the probability of having significant deviations is suppressed unless  $\frac{\Delta_A}{\delta a} \sim \sqrt{d_E}$ .

We can get some intuition on this result by replacing  $\delta a$  with the averaged gap  $\langle \delta a \rangle$  between eigenvalues of the operator  $A$ . In that case, the quotient  $\frac{\Delta_A}{\langle \delta a \rangle} \sim N(A)$  behaves like the number of different outcomes for the operator  $A$ ,  $N(A)$ . We shall return to the importance of the magnitude  $N(A)$  shortly.

**Ensemble vs time averages.** We use here the same strategy as for the ensemble average discussion, but replacing the operator  $A$  by  $\sum_E |E\rangle \langle E| A |E\rangle \langle E|$ . Thus, we want to apply Levy's lemma 3.1 directly. To compute an upper bound for the corresponding Lipschitz constant, we first notice that

$$\langle E | A | E \rangle = \sum_a a |\langle E | a \rangle|^2 \quad \Rightarrow \quad |\langle E | A | E \rangle| \leq \sum_a |a| |\langle E | a \rangle|^2. \quad (3.19)$$

Summing over  $E$ , we reach the conclusion

$$\left| \sum_E |E\rangle \langle E | A | E \rangle \langle E| \right| \leq \|A\|. \quad (3.20)$$

Thus, the norm of the operator  $\sum_E |E\rangle\langle E|A|E\rangle\langle E|$  is bounded above by the norm of the operator  $A$ ,  $\|A\|$ . This means the probability of large deviations satisfies

$$\text{Prob}(|\text{tr}(A\omega_\psi) - \text{tr}(A\Omega_E)| \geq \epsilon) \leq 2e^{-c d_E \epsilon^2 / \|A\|^2}. \quad (3.21)$$

As before, unless the largest eigenvalue of  $A$  scales like  $\sqrt{d_E}$ , the probability that a random time averaged expectation value significantly differs from the ensemble average is exponentially suppressed in the dimension  $d_E$  of  $\mathcal{H}_E$ .

This last statement refers to the notion of quantum ergodicity introduced in [17] (see also the more recent discussion [18]). The latter is based on the equality between time and microcanonical averages. Here, we are saying that for almost every pure state, time average and ensemble average expectation values of operators  $A$  satisfying  $\|A\|^2 \ll d_E$  are equal.

To sum up, the use of typicality allows us to argue that almost every pure state in  $\mathcal{H}_E$  behaves like the ensemble average unless the operator we use to probe it satisfies  $\|A\|^2 \sim d_E$ . Similarly, the fraction of time a random pure state spends away from an equilibrium state is negligible unless  $\frac{\Delta A}{\delta a} \sim N(A) \sim \sqrt{d_E}$ . Furthermore, for almost every pure state, time and microcanonical averages agree when  $\|A\|^2 \ll d_E$ .

### 3.3 Measure of distinguishability

The comparison of expectation values of an observable is not the only way to tell quantum states apart in quantum mechanics. In fact, it is easy to find examples of observables whose measurement can distinguish between different quantum states even if their expectation values are equal. Consider two spin one states  $|0\rangle$  and  $\frac{1}{\sqrt{2}}(|-1\rangle + |1\rangle)$ . By construction, both states have vanishing  $\sigma_z$  expectation values, however a measurement of  $\sigma_z$  can easily tell the states apart.

Notice that in this example we could distinguish both states by comparing the expectation value of  $\sigma_z^2$ . In quantum mechanics, given a state  $\psi$  and an observable  $A$ , the result of its measurement is a set of eigenvalues  $a$  appearing with probabilities  $p_a$ . It is clear that reconstructing the *entire probability spectrum*  $\{p_a\}$  provides more information about the quantum state  $\psi$  than simply measuring  $\langle\psi|A|\psi\rangle$ . In fact, if the observable  $A$  has  $N(A)$  different eigenvalues, we may need the collection of expectation values  $\langle\psi|A^i|\psi\rangle$  for  $i = 1, \dots, N(A)$  to reconstruct such probability spectrum. This reconstruction problem is called the moment problem.

This discussion motivates the notion of distinguishability introduced in the quantum information literature. The *distinguishability* of two quantum states  $\rho$  and  $\sigma$  using a particular observable  $A$  is defined as

$$D_A(\rho, \sigma) = \frac{1}{2} \sum_a |\text{tr}(|a\rangle\langle a|\rho) - \text{tr}(|a\rangle\langle a|\sigma)|, \quad (3.22)$$

where  $|a\rangle$  are the eigenvectors of  $A$ .<sup>6</sup> Notice that this measure is independent of the absolute value of the eigenvalues  $a$  and only depends on the *entire* probability spectrum  $\{p_a\}$ . Thus,

---

<sup>6</sup>In general, the projectors  $|a\rangle\langle a|$  in (3.22) should be replaced by a positive operator valued measure (POVM) [19]. A POVM with a finite number of outcomes is a partition of identity,  $\sum_a M_a = \mathbb{I}$ , where the probability of obtaining  $a$  in the measurement is given by  $\text{tr}(\rho M_a)$ .

it is more appropriate as a measure of quantum state distinguishability than any individual expectation value. In particular, it is shown in [19] that the optimal probability of telling  $\sigma$  and  $\rho$  apart in any measurement is  $\frac{1}{2}(1 + D(\sigma, \rho))$ . This guarantees that if  $D(\rho, \sigma)$  is small, no observable can tell  $\rho$  and  $\sigma$  apart. We stress that the ratio  $\sigma_A^2/(\langle A \rangle)^2$  could go to zero in the limit of large  $d_E$  while  $D_A(\rho, \Omega_E)$  may not.

This notion can be generalized to any set of observables  $\mathcal{A}$ ,

$$D_{\mathcal{A}}(\rho, \sigma) = \max_{A \in \mathcal{A}} D_A(\rho, \sigma) \quad (3.23)$$

irrespectively on whether they commute or not. In particular, if  $\mathcal{A}$  includes the entire set of observables in the Hilbert space, one talks about the *distinguishability* of two quantum states  $\rho$  and  $\sigma$  as,

$$D(\rho, \sigma) = \frac{1}{2} \text{tr} |\rho - \sigma| \quad (3.24)$$

where it is understood the right hand side equals the *maximal* difference in probability spectra achieved over the entire set of available observables [19].

We are interested in computing  $D_A(\psi(t), \Omega_E)$ , that is the distinguishability between a random pure state  $\psi \in \mathcal{H}_E$  and the maximally mixed state  $\Omega_E = \frac{\mathbb{I}_E}{d_E}$ . If one identifies the conditions the observable  $A$  must satisfy for  $D_A(\psi(t), \Omega_E) \rightarrow 0$ , one will conclude the random pure state appears entangled from the perspective of the observable  $A$ . The theorem below summarizes our results.

**Theorem 3.2** *Given a random pure state  $\psi \in \mathcal{H}_E$ , its distinguishability  $D_{\mathcal{A}}(\psi, \Omega_E)$  from the maximally mixed state  $\Omega_E$  using the set of observables  $\mathcal{A}$  satisfies*

$$\begin{aligned} \text{Prob} \left( D_{\mathcal{A}}(\psi, \Omega_E) > \epsilon + \frac{N(\mathcal{A})}{2\sqrt{d_E}} \right) &\leq e^{-\epsilon^2 d_E}, \\ \text{Prob} \left( \langle D_{\mathcal{A}}(\psi(t), \Omega_E) \rangle_t > \epsilon + \frac{N(\mathcal{A})}{2\sqrt{d_E}} \right) &\leq e^{-\epsilon^2 d_E}, \end{aligned} \quad (3.25)$$

for an arbitrary  $\epsilon > 0$ , where  $N(\mathcal{A})$  is the maximum number of outcomes of all measurements in  $\mathcal{A}$ .

The proof is in appendix B.

Since both statements hold for *any*  $\epsilon > 0$ , we learn that the probability for  $D_{\mathcal{A}}(\psi, \Omega_E)$  to be larger than zero is exponentially suppressed in  $d_E$  *unless*  $N(\mathcal{A}) \sim \sqrt{d_E}$ .<sup>7</sup> Furthermore, the second part of the theorem proves that typical states remain indistinguishable from equilibrium for almost all times if  $N(\mathcal{A}) \ll \sqrt{d_E}$ . Notice that both statements hold, in particular, for an individual operator  $A$ . In that case, the number of different outcomes  $N(A)$  refers to that single operator.

Notice theorem 3.2 mathematically characterises the set of operators  $A$  for which a typical random pure state  $\psi \in \mathcal{H}_E$  appears entangled. We can intuitively understand the

---

<sup>7</sup>The theorem is not useful if we consider a set  $\mathcal{A}$  containing of the order of  $\sqrt{d_E}$  different operators with individual different outcomes of order one. But it does not tell us how to implement such measurements in practice.

condition on  $N(\mathcal{A})$  as saying the set of measurements in  $\mathcal{A}$  is not capable of reaching the resolution required to distinguish microstates.

When the entire set  $\mathcal{A}$  has support on a finite subsystem, theorem 3.2 reduces to a rigorous restatement of the well-known result due to Page [20]: the difference between a pure state and a mixed state in a composite Hilbert space is exponentially small unless the number of states being measured is comparable to the square root of the dimensionality of the *entire* Hilbert space. Consider a composed Hilbert space  $\mathcal{H}_S \otimes \mathcal{H}_B$ , where  $\mathcal{H}_S$  and  $\mathcal{H}_B$  are the system and bath Hilbert spaces, respectively. Let  $\mathcal{H}_E \subseteq \mathcal{H}_S \otimes \mathcal{H}_B$  be the subspace associated with the microcanonical ensemble as defined previously. The maximally mixed state in  $\mathcal{H}_E$  is

$$\Omega_E = \frac{\mathbb{I}_E}{d_E}, \quad (3.26)$$

where  $\mathbb{I}_E$  is the identity matrix on  $\mathcal{H}_E$ , whereas its restriction to  $\mathcal{H}_S$  by tracing out over  $\mathcal{H}_B$  defines

$$\Omega_S = \text{tr}_B(\Omega_E). \quad (3.27)$$

Then, from theorem 3.2 we obtain

$$\text{Prob} \left( D_{\mathcal{A}}(\psi, \Omega_E) > \epsilon + \frac{d_S}{2\sqrt{d_E}} \right) \leq e^{-c\epsilon^2 d_E}, \quad (3.28)$$

since the number of outcomes is necessarily bounded by the dimension  $d_S$  of the Hilbert space  $\mathcal{H}_S$ . The last statement is equivalent to the bound found in [2, 3]

$$\langle \|\rho_S - \Omega_S\| \rangle_S \leq \sqrt{\frac{d_S^2}{d_E}}. \quad (3.29)$$

An equivalent, perhaps more explicit, way of reproducing Page's original discussion is by considering a random quantum state for a system of  $N$  spins  $\sigma_i = \pm \frac{1}{2}$

$$|\psi\rangle = \sum_{\{\sigma_i\}} \frac{1}{2^{N/2}} |\sigma_1, \dots, \sigma_N\rangle e^{i\theta(\sigma_1, \dots, \sigma_N)} \quad (3.30)$$

where  $\theta(\sigma_1, \dots, \sigma_N)$  is a set of random phases. If only  $k$  spins are measured, any observable acting on that subspace of the Hilbert space  $\mathcal{H}_S$  will have a number of outcomes bounded above from its dimensionality  $d_S = 2^k$ . Its expectation values can be evaluated using the density matrix  $\rho_k$ . Explicit calculation gives [5, 20]

$$\rho_k = \frac{1}{2^k} \sum_{\{\sigma_k\}} |\sigma_1, \dots, \sigma_k\rangle \langle \sigma_1, \dots, \sigma_k| + \mathcal{O}(2^{-N+2k}) = \Omega_k + \mathcal{O}(2^{-N+2k}). \quad (3.31)$$

Thus,  $\rho_k$  is well approximated by the maximally entangled mixed state  $\Omega_k$  unless  $k \sim \frac{N}{2}$ , which is equivalent to the requirement above  $d_S^2 \sim 2^N = d_E$ .

This matches the content of Page's initial result [20], emphasizing the role played by quantum entanglement and concentration of measure [13, 14], which is the mathematical justification behind typicality.

### 3.4 Coarse-grained observables

Theorem 3.2 is relevant for first principle considerations, but it does not say whether such distinguishing measurements can be performed. For example, one could conceive the existence of operators whose spectra satisfy the requirement  $N(A) \sim \sqrt{d_E}$  but have an averaged eigenvalue difference suppressed in  $d_E$  itself. Since obtaining information about a system is never for free in quantum mechanics [21], it is an important question to discuss whether such measurements can be implemented in short time in the probe limit.

The uncertainty principle already teaches us that achieving high precision (large  $N(A)$ ) in a measurement performed in short time requires large energies. In this section we show that any measurement performed with finite time and energy resources implements the notion of a *coarse-grained observable* [17].

It is instructive to review the standard approach to measurements in field theory. Given an observable  $\phi(x, t)$ , its measurement is described in terms of a semiclassical process in which a classical source  $J$  couples to  $\phi(x, t)$  by turning on an interaction  $H_{\text{int}} = \lambda \phi J \delta(x)$  at time  $t$ , and turning it off after an infinitesimally short amount of time. In order to keep the disturbance minimal (probe limit), one takes the limit  $J \rightarrow 0$ . A semi-classical measurement is not limited in precision, and correlation functions found this way are arbitrarily accurate. However, this treatment only holds approximately as a limit of a quantum process that we now describe.

To describe a quantum measurement, we follow von Neumann [22]. We entangle our physical system with a quantum detector and consider a sharp projection on the compound state after the measurement takes place. More mathematically, assume our initial state is in a product state  $|\psi\rangle \otimes |\alpha\rangle \in \mathcal{H}_E \otimes \mathcal{H}_A$ , where  $\mathcal{H}_A$  stands for the apparatus Hilbert space. We turn on an interaction Hamiltonian  $H_{\text{int}} = \lambda A \otimes J$ , where  $A$  is the observable we want to measure, and  $J$  acts on  $\mathcal{H}_A$ .

As before the interaction acts at a time scale  $\epsilon$  much smaller than the inverse energy of the system such that the evolution due to the physical system Hamiltonian during the measurement can be ignored. The analogue of the probe limit is the constraint that the kick received by the ensemble due to the measurements is small, i.e.  $S_{\text{int}}/\hbar = \int dt H_{\text{int}}/\hbar \ll E\tau/\hbar$  where  $\tau = \mathcal{O}(1)$ .

Let us expand the time evolution of the initial entangled state in the eigenbasis of  $A$

$$|\Psi(t)\rangle = \sum_a \psi_a(t) e^{-ia\lambda J t} |a\rangle |\alpha\rangle. \quad (3.32)$$

The measurement will distinguish the outcome  $a$  from  $a + \delta a$  if the apparatus wave-functions corresponding to these different eigenvalues become orthogonal at time  $t$

$$S(t) = \langle \alpha | e^{-i\lambda(\delta a)J t} | \alpha \rangle = 0. \quad (3.33)$$

This condition relates the amount of time and energy involved in a quantum measurement with resolution  $\delta a$  in our observable  $A$ . It allows us to find a lower bound on the smallest resolvable eigenvalue gap  $\delta a$ .

**Theorem 3.3** *The smallest gap  $\delta a$  resolvable in a measurement of an observable  $A$  in time  $t$  using the interaction Hamiltonian  $H_{\text{int}} = \lambda A \otimes J$  is bounded below by*

$$\frac{\pi \hbar \langle A \rangle}{\langle H_{\text{int}} \rangle t} = \frac{\pi \hbar}{\lambda \langle J \rangle t} \leq \delta a. \quad (3.34)$$

The proof appears in appendix C, and it is based on the existence of a universal limit on how fast a state can dynamically evolve to an orthogonal state in an isolated quantum mechanical system [23].

Using (3.34), we can derive the relation

$$N(A) \leq \frac{\Delta_A}{\delta a} \leq \frac{\Delta_A}{\langle A \rangle} \frac{\langle H_{\text{int}} \rangle t}{\pi \hbar}. \quad (3.35)$$

Measurements that distinguish microstates, i.e.  $N(A) \sim \sqrt{d_E}$  always require a large kick. Thus, they always involve long times or large energy resources.<sup>8</sup>

**Relation to coarse-grained observables.** We want to show that if the measurement action  $S_{\text{int}}/\hbar$  is *finite*, i.e. if the amount of time and energy resources are finite, the fine grained observable  $A$  effectively behaves as a coarse-grained observable due to the finite precision  $\delta a$  achievable by the measurement. Indeed, given a precision  $\delta a$ , any observable  $A$  allows a description in terms of a coarse-grained observable

$$\tilde{A} = \sum_{i=0}^{N(\tilde{A})} \tilde{a}_i \Pi_i, \quad \tilde{a}_i = \tilde{a}_0 + i \delta a, \quad \Pi_i = \sum_{a \in [\tilde{a}_i, \tilde{a}_{i+1})} |a\rangle \langle a|. \quad (3.36)$$

Its eigenvalues  $\tilde{a}_i$  are defined as follows:  $\tilde{a}_0 = n_0 \delta a$ , where  $n_0 \delta a \leq a_{\min} < (n_0 + 1) \delta a$ , whereas  $n_\star \delta a \leq a_{\max} < (n_\star + 1) \delta a$ . Thus, the number of coarse-grained eigenvalues  $N(\tilde{A})$  equals  $n_\star - n_0 + 1$  and the projectors  $\Pi_i$  include *all* the microscopic eigenvalues  $a$  in the macroscopic eigen-band  $(n_0 + i) \delta a \leq a < (n_0 + i + 1) \delta a$ .

Finite action  $S_{\text{int}}/\hbar$  requires  $\langle A \rangle \sim \delta a$ . This is equivalent to  $n_0, n_\star \sim \mathcal{O}(1)$ . Equivalently, even if the observable  $A$  has no degenerate eigenvalues, i.e. the number of microscopic outcomes is  $d_E$ , the number of macroscopic outcomes attainable with finite precision is order one, i.e. the observable  $A$  behaves like a coarse-grained observable  $\tilde{A}$ .<sup>9</sup>

## 4 Lessons for quantum gravity

In this section we discuss the implications of our results for black hole physics in an AdS/CFT context [25, 26].<sup>10</sup> Even though our comments are far more general, we shall use

<sup>8</sup>Note that for measurements that take a long time the self-Hamiltonian of the system cannot be ignored, and acts as a source of noise.

<sup>9</sup>The connection between information loss in black holes and finite resolution of low energy measurements has been previously discussed in [24].

<sup>10</sup>Any precise application of our results to any quantum theory of gravity would require to extend our analysis not only to infinite dimensional Hilbert spaces but also to *unbounded* operators. The techniques we used here to prove equilibration using typicality have been generalized to infinite systems in [27]. Even though we believe that an understanding and mathematical formulation of this extension for QFTs and CFTs is of great importance, we shall not pursue this here.

the language of  $\mathcal{N} = 4$  SYM with gauge group  $SU(N)$  and its holographic dual, when appropriate. Large  $AdS_5$  black holes have both energies and entropies scaling as  $N^2$ ,  $S \sim E$ . Hence, the dimension of the microcanonical ensemble  $d_E$  is exponentially large in the energy of the microstates, i.e.  $d_E \sim e^{aE}$ , where  $a$  is an order one number.

#### 4.1 Breaking down of effective field theory

Our analysis finds that any quantum measurement capable of resolving the differences between random pure states and a maximally entangled one in large dimensional Hilbert spaces requires an exponentially large amount of resources, i.e. either energies or times exponential in  $N^2$ . Clearly, any low energy effective action does not include such degrees of freedom. For example, in AdS/CFT, the supergravity approximation deals with operators with conformal dimension  $\Delta \sim \mathcal{O}(1)$  (no scaling in  $N$ ), whereas classical supergravity saddle points have energies of order  $N^2 \sim \log d_E$ . Thus, we must conclude that *all low energy observables* in such effective theory can only achieve this task by waiting  $\log t = \mathcal{O}(N^2)$ .<sup>11</sup> This is in agreement with the finding that 2-point correlators of light operators detect deviations from thermality at these time-scales [7].

Equivalently, *random pure states appear entangled for all semiclassical gravitational probes in time polynomial in  $N$* . We emphasize this statement excludes those operators built out of a large number of products of light operators. These are heavy and can distinguish microstates. Thus, these operators of the full quantum theory do *not* belong to the effective theory, i.e. the set of semiclassical operators  $\mathcal{A}_{\text{low}}$  does *not* form an algebra. If one considers operators made out of the product of order  $N^2$  light observables, one expects perturbation theory to break down due to the large combinatorial factors appearing in Feynman diagrams, similar to the appearance of non-planar effects in non-abelian gauge theories [28].<sup>12</sup>

Both, our quantum mechanical results in section 3 and the arguments above suggest that the number of operator insertions,  $N$  in equation (2.5) and their conformal dimensions  $\Delta_i = \Delta(\mathcal{O}_i)$  play a similar role to the number of outcomes in our quantum mechanics discussion [5, 6, 9, 29–31]. We postpone a precise mathematical formulation of this problem to future work. But we stress this expectation holds in effective field theory. This is because there exists a map between low energy operators and particle excitations in this regime. Thus, the number of operator insertions in a given correlator describes how large the dimension of the perturbative Fock space is being explored. In this respect, the argument is analogous to our spin discussion in section 3.3.

<sup>11</sup>Even though it is hard to prove by first principles, it is believed that any high precision measurement, as the ones we require in our discussion, lies beyond the regime of validity of an effective field theory. We thank Zohar Komargodski for emphasising this to us.

<sup>12</sup>Perturbation theory in gravity is an expansion in Newton's constant  $G_N$ . When computing correlators in black hole microstates, the perturbative expansion is expected to be in  $1/S$ , where  $S = A/(4G_N)$  is the standard Hawking-Bekenstein formula. For correlators involving  $m \sim N^2 \sim S$  insertions, one expects such perturbative arguments to break down.



## 4.2 Effective emergence of locality

Holography allows us to apply our quantum mechanical results to gravitational physics. In particular, we know the low energy sector of the bulk physics involves classical gravity, where *locality* is manifest. Our work proves that any operator capable of distinguishing a random pure state from a mixed state does not belong to the low energy effective action. Thus, inverting the logic presented in the last subsection, we can say that the notion of bulk locality emerges when we restrict the entire set of observables to  $\mathcal{A}_{\text{low}}$ .

We finish this work emphasizing that the restriction to the subspace of observables compatible with an effective notion of bulk locality is analogous to the recent proposal by Papadodimas & Raju [30, 31] reconstructing *local bulk operators* in the interior of a black hole from boundary quantum data, building on the seminal work in [32–36].

**Emergence of bulk locality and relation to Papadodimas & Raju.** According to [32–36], local bulk operators  $\phi_{\text{CFT}}(t, \Omega, z)$  in the exterior of a black hole can be constructed as

$$\phi_{\text{CFT}}^i(t, \Omega, z) = \sum_{\vec{\ell}, \omega > 0} (\mathcal{O}_{\vec{\ell}, \omega}^i f_{\vec{\ell}, \omega}(t, \Omega, z) + \text{h.c.}) \quad (4.1)$$

where  $(t, \Omega, z)$  are boundary labels which can be interpreted as bulk AdS coordinates,  $\mathcal{O}_{\vec{\ell}, \omega}^i$  are the Fourier modes of a boundary local operator on the sphere and  $f_{\vec{\ell}, \omega}(t, \Omega, z)$  are appropriately chosen functions. Such bulk operators can be constructed, order by order in a  $\frac{1}{N}$  expansion [37].

Recently, this construction was claimed to be extended for operators probing the *interior* of a black hole [30, 31]. In this case, the field (4.1) is replaced by the mode expansion

$$\phi_{\text{CFT}}^i(t, \Omega, z) = \sum_{\vec{\ell}, \omega > 0} \left( \mathcal{O}_{\vec{\ell}, \omega}^i g_{\vec{\ell}, \omega}^{(1)}(t, \Omega, z) + \tilde{\mathcal{O}}_{\vec{\ell}, \omega}^i g_{\vec{\ell}, \omega}^{(2)}(t, \Omega, z) \right) \quad (4.2)$$

where  $g_{\vec{\ell}, \omega}^{(a)}(t, \Omega, z)$   $a = 1, 2$  can be found in [38]. The expansion (4.2) involves two sets of operator modes:  $\mathcal{O}_{\vec{\ell}, \omega}^i$  as before and the *mirror operators*  $\tilde{\mathcal{O}}_{\vec{\ell}, \omega}^i$ . The latter were defined in [30, 31] as those satisfying

$$\tilde{\mathcal{O}}_{\vec{\ell}, \omega}^i A_\alpha |\psi\rangle = A_\alpha e^{-\beta\omega} \mathcal{O}_{-\vec{\ell}, -\omega}^i |\psi\rangle \quad (4.3)$$

where  $A_\alpha$  belongs to the subset of local boundary observables

$$A_\alpha = \sum_P \alpha(P) (\mathcal{O}_{\vec{\ell}, \omega}^i)^{P(i, \omega, \vec{\ell})} \quad (4.4)$$

belonging to the low energy sector  $\mathcal{A}_{\text{low}}$ . This last condition was more precisely stated in [30, 31] by requiring the functions  $P(i, \omega, \vec{\ell})$  to satisfy

$$\sum_{i, \vec{\ell}, \omega} P(i, \omega, \vec{\ell}) \omega \ll N \quad (4.5)$$



Thus, by construction,  $A_\alpha \in \mathcal{A}_{\text{low}}$  in our previous discussion. Papadodimas & Raju showed mirror operators exist and that they depend on the specific state  $|\psi\rangle$  where they act on [30, 31].

The defining equation (4.3) guarantees that boundary correlators are compatible with thermal behaviour. Furthermore, all operators  $A_\alpha \in \mathcal{A}_{\text{low}}$  satisfy  $A_\alpha|\psi\rangle \neq 0$ . Thus, states  $|\psi\rangle$  do appear entangled for the subset of observables  $A_\alpha$  satisfying (4.5).

The connection with our work is as follows. We characterised the set of operators that do not distinguish random pure states from the maximally entangled state  $\Omega_E$  in a microcanonical ensemble  $\mathcal{H}_E$ . Our analysis was in the entire quantum theory and our only assumption on the spectrum was the existence of non-degenerate gaps to exclude integrable systems.<sup>13</sup> When we embed our discussion into a holographic theory, the latter is believed to have an spectrum of relatively sparsely  $\mathcal{O}(1)$  excitations separated by a large gap from a dense set of heavy states (black hole microstates) [4]. In these theories, it is natural to identify the subset of operators  $\mathcal{A}_{\text{low}}$  with those satisfying the more precise constraint (4.5). These satisfy

$$D_{A_\alpha}(\psi, \Omega_E) \rightarrow 0 \quad \forall A_\alpha \in \mathcal{A}_{\text{low}} \quad (4.6)$$

Thus all operators considered in [30, 31] are expected to satisfy (4.6).

These statements can be made more explicit for a system of  $N$  spins. If we restrict ourselves to the subset of observables acting on subsystems of  $k$  spins, where  $k \ll \frac{N}{2}$ , we already showed in section 3 that random pure states appear entangled. Thus, in this example it is clear that the set of operators  $A_\alpha$  to be considered are those satisfying the property (4.6).

A further outcome of this discussion is that the construction in [30, 31] may be less state dependent than what it appears, since all operators  $A_\alpha$  see typical random pure states  $|\psi\rangle \in \mathcal{H}_E$  as  $\Omega_E$ . The latter follows from typicality and as such, it is necessarily a probabilistic statement at this stage that requires further investigation.

Having discussed the connection between our results and the emergence of locality, we comment below on the implications for complementarity and the existence of a black hole interior. Some of the considerations below are necessarily similar to the ones also discussed in [30, 31].

**Complementarity.** One of the main conceptual issues in black hole physics regards the compatibility of black hole complementarity [39, 40] with the preservation of bulk locality. One consequence of the former is that the degrees of freedom inside and outside of the black hole are *not* independent. The possible tension between both concepts can easily be phrased. Low energy bulk observers using an EFT description would expect correlation functions between bulk operators in the interior ( $z_{\text{int}}$ ) and exterior ( $z_{\text{ext}}$ ) of the black hole to vanish

$$[\phi_{\text{EFT}}(t, \Omega, z_{\text{int}}), \phi'_{\text{EFT}}(t', \Omega', z_{\text{ext}})]_{\text{EFT}} = 0. \quad (4.7)$$

This is because in the EFT approximation, one is doing a quantum field theory calculation in a black hole background and the points  $z_{\text{int}}$  and  $z_{\text{ext}}$  are causally disconnected. From the

---

<sup>13</sup>The non-degenerate gap assumption can be further relaxed [10].

exact quantum theory of gravity perspective, one is probing the operator equation (4.7) in its heavy sector using the CFT operators (4.2). That is, one is considering correlators of the form

$$\Phi_i(z_{\text{int}}, z_{\text{ext}}) \equiv \langle \psi_i | [\phi_{\text{CFT}}(t, \Omega, z_{\text{int}}), \phi'_{\text{CFT}}(t', \Omega', z_{\text{ext}})] | \psi_i \rangle \quad (4.8)$$

with  $\psi_i \in \mathcal{H}_E$  being the black hole microstates. One expects that such correlators are well approximated by the thermal correlator. Consequently,

$$\begin{aligned} \Phi_i(z_{\text{int}}, z_{\text{ext}}) &\approx \langle \psi_{\text{BH}} | [\phi_{\text{CFT}}(t, \Omega, z_{\text{int}}), \phi'_{\text{CFT}}(t', \Omega', z_{\text{ext}})] | \psi_{\text{BH}} \rangle \\ &\approx [\phi_{\text{EFT}}(t, \Omega, z_{\text{int}}), \phi'_{\text{EFT}}(t', \Omega', z_{\text{ext}})]_{\text{EFT}} = 0. \end{aligned} \quad (4.9)$$

This would seem to violate complementarity given the expected non-independence of interior and exterior black hole degrees of freedom.

Our results in sections 3.3 and 3.4 provide mathematical evidence for the existence of a conceptual framework to understand in which sense both statements can be compatible. By first principles, the commutator (4.8) is *non-vanishing*. For the subset of operators in  $\mathcal{A}_{\text{low}}$  for which an effective local bulk description exists, deviations from thermal behaviour are exponentially suppressed in the entropy of the black hole

$$\Phi_i(z_{\text{int}}, z_{\text{ext}}) = \mathcal{O}(e^{-S_{\text{BH}}}) \quad \forall \phi_{\text{CFT}}, \phi'_{\text{CFT}} \in \mathcal{A}_{\text{low}} \quad (4.10)$$

Thus, even though these corrections are difficult to measure, particularly so in any semiclassical approximation, they are still responsible for the non-vanishing of the exact quantum gravity answer. For the subset of operators for which no notion of locality exists, there is no reason to expect the commutator to vanish. Thus, there was no tension to begin with.

In quantum mechanics, it is possible to construct *commuting coarse-grained* operators from microscopic non-commuting ones, starting with the momentum and position operators [17], explaining why classical apparatus can measure them simultaneously. What we are saying here is that not only low energy gravity probes appear to be entangled, but they may also appear to commute due to their coarse-grained nature, even though this is not true for the exact microscopic operators.

**Existence of a geometric black hole interior.** There are two main reasons why we associate black holes with thermal states in the AdS/CFT correspondence. First, because in its euclidean path integral formulation, the euclidean AdS black hole is the dominant saddle contribution (at large enough temperatures) and matches the CFT expectation values of a thermal density matrix [41]. Second, because eternal AdS black holes are believed to be described by the entangled state [7]

$$\rho = |\psi_{\text{BH}}\rangle\langle\psi_{\text{BH}}|, \quad \text{with} \quad |\psi_{\text{BH}}\rangle = \sum_{n \in \mathcal{H}} \frac{e^{-\beta E_n/2}}{\sqrt{Z}} |n\rangle_{\mathcal{H}_L} \otimes |n\rangle_{\mathcal{H}_R} \quad (4.11)$$

where  $\mathcal{H}_L$  and  $\mathcal{H}_R$  stand for the two isomorphic Hilbert spaces defined on each of their two conformal boundaries. An asymptotic observer living on  $\mathcal{H}_R$  will only measure observables

$\mathcal{O}_R$  acting on  $\mathcal{H}_R$ . By construction, such expectation values are captured by the reduced density matrix obtained by tracing over  $\mathcal{H}_L$

$$\rho_R = \text{tr}_{\mathcal{H}_L} \rho = \frac{1}{Z} \sum_{n \in \mathcal{H}_R} e^{-\beta E_n} |n\rangle \langle n|. \quad (4.12)$$

Thus, these expectation values are manifestly thermal

$$\text{tr}_{\mathcal{H}_L \otimes \mathcal{H}_R} (\rho \mathcal{O}_R) = \text{tr}_{\mathcal{H}_R} (\rho_R \mathcal{O}_R) = \frac{1}{Z} \sum_{n \in \mathcal{H}_R} e^{-\beta E_n} \langle n | \mathcal{O}_R | n \rangle. \quad (4.13)$$

One natural question this eternal AdS black hole description raises is: what is the holographic dual of the quantum state  $\rho_R$ ? The reason to ask this question is because there are many different quantum states in  $\mathcal{H}_L \otimes \mathcal{H}_R$  giving rise to the same density matrix  $\rho_R$  capturing the physics of a single asymptotic observer. This suggests that  $\rho_R$  only describes the exterior of the AdS event horizon [42–45].

In fact, this is also heuristically suggested by euclidean path integral considerations. The euclidean black hole is a smooth geometry constructed from the lorentzian geometry by cutting the horizon, *removing its interior* and gluing it in a smooth way. It is indeed true that such euclidean saddle knows about the full entropy of the system, but it does not know anything about the specific microstates of the system [46].<sup>14</sup> This perspective suggests that information about the black hole microstates is “hidden” in the interior, but does such interior allow for a local bulk description?

The arguments used to reconcile complementarity with EFT locality expectations already suggest the answer depends on the observables one studies. Whenever the effective black hole geometry is reliable, that is when we probe it with operators  $\phi_{\text{CFT}} \in \mathcal{A}_{\text{low}}$ , we can obviously not distinguish between the black hole and the microstates. There should exist some effective notion of interior. This is made more precise in [30, 31]. If an asymptotic observer attempts to improve on this situation, he/she must either wait for a long time or consider correlators involving operators *not* in  $\mathcal{A}_{\text{low}}$ . These operators can carry very large energies, could distinguish among black hole microstates and will generically have large variances in the ensemble of microstates. Thus, there is no well defined notion of locality for these.

## Acknowledgments

We would like to thank Vijay Balasubramanian, José Barbón, Micha Berkooz, Ignacio Cirac, Patrick Hayden, Seth Lloyd, Simon Ross, James Sully, and Leonard Susskind for discussions. JS would like to thank University of Oviedo and the organizers and participants of the *Quantum Information* workshop in Benasque, *Black holes in String Theory* workshop in Ann Arbor and the COST-WIS workshop on *Black holes and quantum information* in the Weizmann Institute for hospitality during different stages of this work. The work of NL is

<sup>14</sup>For a recent discussion on the relation of microstates to thermodynamics and euclidean path integral considerations, see [47].

supported in part by the Natural Sciences and Engineering Research Council of Canada. The work of JS was partially supported by the Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/G007985/1], the Science and Technology Facilities Council (STFC) [grant number ST/J000329/1] and a *Visiting Professor* fellowship at the Universidad de Oviedo.

## A Ensembles, averages & variances

We briefly review the definition of the ensemble and time averages discussed in section 3.1.

**Ensemble averages.** The most general parameterisation of the ensemble of states  $\psi \in \mathcal{H}_E$  is in terms of the arbitrary set of coefficients  $c_E$  describing the expansion of *any* state  $\psi$  in the eigen-energy basis.

$$|\psi(0)\rangle = \sum_E c_E^\psi |E\rangle \quad \text{with} \quad \sum_E |c_E^\psi|^2 = 1 \quad (\text{A.1})$$

Given an observable  $\mathcal{O}$ , its expectation value on such a random state equals.

$$\langle \psi(0) | \mathcal{O} | \psi(0) \rangle = \sum_{E, E'} c_E^\psi c_{E'}^{\psi*} \langle E' | \mathcal{O} | E \rangle \quad (\text{A.2})$$

The averaged expectation value  $\langle \mathcal{O} \rangle_E$  is defined as,

$$\langle \mathcal{O} \rangle_E = \int d\vec{c}^\psi \sum_{E, E'} c_E^\psi c_{E'}^{\psi*} \langle E' | \mathcal{O} | E \rangle \quad (\text{A.3})$$

where the measure is normalised to one, i.e.  $\int d\vec{c}^\psi = 1$ .

This measure satisfies the property [8]

$$\int d\vec{c}^\psi c_E^\psi c_{E'}^{\psi*} = \frac{1}{d_E} \delta_{EE'}. \quad (\text{A.4})$$

Equivalently, the average dephases the state. Using (A.4) in (A.3),

$$\langle \mathcal{O} \rangle_E = \frac{1}{d_E} \text{tr } \mathcal{O} = \text{tr}(\mathcal{O} \Omega_E) \quad \text{with} \quad \Omega_E = \frac{\mathbb{I}_E}{d_E} \quad (\text{A.5})$$

reproducing the statement (3.2) in the main text.

We define the variance  $\sigma_{\mathcal{O}}^2$  as for any other ensemble.

$$\sigma_{\mathcal{O}}^2 \equiv \langle (\langle \psi | \mathcal{O} | \psi \rangle - \langle \mathcal{O} \rangle_E)^2 \rangle_E \quad (\text{A.6})$$

This satisfies  $\sigma_{\mathcal{O}}^2 = \langle \langle \psi | \mathcal{O} | \psi \rangle^2 \rangle_E - \frac{(\text{tr } \mathcal{O})^2}{d_E^2}$  which equals

$$\sigma_{\mathcal{O}}^2 = \int d\vec{c}^\psi \left( \sum_{E_1, E_2, E_3, E_4} c_{E_1}^\psi c_{E_2}^{\psi*} c_{E_3}^\psi c_{E_4}^{\psi*} \langle E_2 | \mathcal{O} | E_1 \rangle \langle E_3 | \mathcal{O} | E_4 \rangle \right) - \frac{(\text{tr } \mathcal{O})^2}{d_E^2} \quad (\text{A.7})$$

Using the integral identity [8]

$$\int d\vec{c}^\psi |c_E^\psi|^2 |c_{E'}^\psi|^2 = \frac{1 + \delta_{EE'}}{d_E(d_E + 1)}, \quad (\text{A.8})$$

in the first term and taking into account that the same dephasing as in (A.4) occurs, we conclude

$$\sigma_{\mathcal{O}}^2 = \frac{1}{d_E + 1} \left( \frac{\text{tr}(\mathcal{O}^2)}{d_E} - \frac{\text{tr}(\mathcal{O})^2}{d_E^2} \right). \quad (\text{A.9})$$

There exists a different commonly used possibility to describe ensemble averages using unitary integrals. In this case, one averages the expectation values of a given operator  $\mathcal{O}$  over the entire set of states obtained from  $\psi$  by a unitary transformation  $U$ . The averaged expectation value equals

$$\langle \mathcal{O} \rangle_\psi = \int dU \langle \psi \otimes U | \mathcal{O} | U \otimes \psi \rangle = \frac{1}{d_E} \text{tr } \mathcal{O} \quad (\text{A.10})$$

where we used the identity [48]

$$\int dU U_{ij} U_{kl}^* = \frac{1}{d_E} \delta_{ik} \delta_{jl}. \quad (\text{A.11})$$

Thus, both ensemble averages are equal and match the microcanonical average by construction.

The variance is defined as before, but using the unitary integral average

$$\sigma_{\mathcal{O}}^2 \equiv \langle (\langle \psi | \mathcal{O} | \psi \rangle - \text{tr}(\mathcal{O} \Omega_E))^2 \rangle_\psi. \quad (\text{A.12})$$

The only non-trivial calculation left is

$$\begin{aligned} \langle \langle \psi | \mathcal{O} | \psi \rangle^2 \rangle_\psi &= \int dU \langle \psi \otimes \psi | U \otimes U (\mathcal{O} \otimes \mathcal{O}) U^\dagger \otimes U^\dagger | \psi \otimes \psi \rangle \\ &= \frac{1}{d_E^2 - 1} \langle \psi \otimes \psi | \left( (\text{tr } \mathcal{O})^2 - \frac{1}{d_E} \text{tr}(\mathcal{O}^2) \right) \mathbb{I} \\ &\quad + \left( \text{tr}(\mathcal{O}^2) - \frac{1}{d_E} (\text{tr } \mathcal{O})^2 \right) \mathbb{S} | \psi \otimes \psi \rangle \\ &= \frac{1}{d_E(d_E + 1)} [\text{tr}(\mathcal{O}^2) + (\text{tr } \mathcal{O})^2] \end{aligned} \quad (\text{A.13})$$

where we used the identity [48]

$$\begin{aligned} \int dU U_{i_1 j_1} U_{i_2 j_2} U_{i'_1 j'_1}^* U_{i'_2 j'_2}^* &= \frac{1}{d_E^2 - 1} (\delta_{i_1 i'_1} \delta_{i_2 i'_2} \delta_{j_1 j'_1} \delta_{j_2 j'_2} + \delta_{i_1 i'_2} \delta_{i_2 i'_1} \delta_{j_1 j'_2} \delta_{j_2 j'_1}) \\ &\quad - \frac{1}{d_E(d_E - 1)} (\delta_{i_1 i'_1} \delta_{i_2 i'_2} \delta_{j_1 j'_2} \delta_{j_2 j'_1} + \delta_{i_1 i'_2} \delta_{i_2 i'_1} \delta_{j_1 j'_1} \delta_{j_2 j'_2}) \end{aligned} \quad (\text{A.14})$$

and the definition of the swap operator  $\mathbb{S}$  as  $\mathbb{S}|i, j\rangle = |j, i\rangle$ . Plugging this back into (A.12), we reproduce the variance (A.9) computed above.

**Time averages.** In this ensemble, we consider the time evolution of the most general pure state  $\psi \in \mathcal{H}_E$  in (A.1)

$$|\psi(t)\rangle = \sum_E c_E^\psi e^{-iEt} |E\rangle, \quad (\text{A.15})$$

viewing any expectation value  $\langle \psi(t) | \mathcal{O} | \psi(t) \rangle$  as a random variable uniformly distributed over  $t \in [0, \infty)$ , so that the time averaged expectation value is defined by

$$\langle \mathcal{O} \rangle_t = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt \langle \psi(t) | \mathcal{O} | \psi(t) \rangle = \sum_E |c_E^\psi|^2 \langle E | \mathcal{O} | E \rangle = \text{tr}(\mathcal{O} \omega_\psi), \quad (\text{A.16})$$

where

$$\omega_\psi \equiv \langle \psi(t) \rangle_t = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt |\psi(t)\rangle \langle \psi(t)| = \sum_E |c_E|^2 |E\rangle \langle E| \quad (\text{A.17})$$

is the time averaged of the density matrix  $\rho^\psi(t) = |\psi(t)\rangle \langle \psi(t)|$  constructed out of the initial pure state (A.15).

The variance is defined as in any other ensemble as

$$\sigma_{\mathcal{O}}^2(\psi) = \langle \left( \text{tr}(\mathcal{O} \psi(t)) - \langle \mathcal{O} \rangle_t \right)^2 \rangle_t. \quad (\text{A.18})$$

Since  $\langle \mathcal{O} \rangle_t$  is diagonal, this variance only has contributions from the off-diagonal matrix elements of the operator  $\mathcal{O}$  in the energy eigenbasis, so that

$$\begin{aligned} \sigma_{\mathcal{O}}^2(\psi) &= \left\langle \left( \sum_{E_1, E_2} c_{E_1} c_{E_2}^* e^{i(E_2 - E_1)t} \langle E_2 | \mathcal{O} | E_1 \rangle \right)^2 \right\rangle_t \\ &= \sum_{E_1 \neq E_2, E_3 \neq E_4} c_{E_1} c_{E_2}^* c_{E_3} c_{E_4}^* \langle e^{i(E_2 - E_1 + E_4 - E_3)t} \rangle_t \langle E_2 | \mathcal{O} | E_1 \rangle \langle E_4 | \mathcal{O} | E_3 \rangle. \end{aligned} \quad (\text{A.19})$$

The time average forces the exponent  $E_2 - E_1 + E_4 - E_3$  to vanish. It is here where the assumption on the Hamiltonian having non-degenerate energy gaps enters. Assuming the latter and using that  $E_1 \neq E_2$  and  $E_3 \neq E_4$ , we conclude the only terms surviving the intrinsic dephasing due to the time average are  $E_1 = E_4$  and  $E_2 = E_3$ , so that the variance equals

$$\sigma_{\mathcal{O}}^2(\psi) = \sum_{E, E'} |c_E|^2 |c_{E'}|^2 |\langle E' | \mathcal{O} | E \rangle|^2 - \sum_E |c_E|^4 |\langle E | \mathcal{O} | E \rangle|^2, \quad (\text{A.20})$$

where we added and subtracted the last term to write the final more symmetric expression.

## B Proof of theorem 3.2

As most statements regarding probabilities in Hilbert spaces of large dimensionality, the strategy of the proof relies in the use of Levy's lemma 3.1. First, choose the function  $f(\psi)$  in Levy's lemma to be the distinguishability of  $\psi \in \mathcal{H}_E$  from the maximally mixed state  $\Omega_E = \mathbb{I}_E/d_E$  using the set of observables  $\mathcal{A}$ ,  $D_{\mathcal{A}}(\psi, \Omega_E)$ .

We need to compute the ensemble average  $\langle D_{\mathcal{A}}(\psi, \Omega_E) \rangle_\psi$ . Due to (3.23), we can bound this average by.

$$\langle D_{\mathcal{A}}(\psi, \Omega_E) \rangle_\psi \leq \sum_{A \in \mathcal{A}} \langle D_A(\psi, \Omega_E) \rangle_\psi \quad (\text{B.1})$$

This allows us to restrict the application of Levy's lemma to a single observable  $A \in \mathcal{A}$ . Thus, we need to compute  $\langle D_A(\psi, \Omega_E) \rangle_\psi$  and its associated Lipschitz constant  $\lambda$ .

Since  $\|D_A(\psi, \Omega_E)\| \leq 1$ , because each individual state probability is bounded by unity, we conclude using (3.15) that  $\lambda \leq 2$ . Next, we find an upper bound on  $\langle D_A(\psi, \Omega_E) \rangle_\psi$  where the observable  $A \in \mathcal{A}$  is described by POVM elements  $M_a$ .

$$\begin{aligned}
 \langle D_A(\psi, \Omega_E) \rangle_\psi &= \frac{1}{2} \sum_a \langle |\text{tr}(M_a \psi) - \text{tr}(M_a \Omega_E)| \rangle_\psi \\
 &= \frac{1}{2} \sum_a \langle \sqrt{\text{tr}(M_a^E (\psi - \Omega_E))^2} \rangle_\psi \\
 &\leq \frac{1}{2} \sum_a \sqrt{\langle \text{tr}(M_a^E (\psi - \Omega_E))^2 \rangle_\psi} \\
 &\leq \frac{1}{2} \sum_a \sqrt{\langle \text{tr}(M_a^E \psi)^2 \rangle_\psi - \text{tr}(M_a^E \Omega_E)^2},
 \end{aligned} \tag{B.2}$$

$M_a^E = \Pi^E M_a \Pi^E$  is the restriction of  $M_a$  to  $\mathcal{H}_E$ , we used the Cauchy-Schwarz inequality in the third line and used (A.10) to write  $\langle \psi \rangle_\psi = \Omega_E$ .<sup>15</sup> Using (A.9),

$$\begin{aligned}
 \langle D_A(\psi, \Omega_E) \rangle_\psi &\leq \frac{1}{2} \sum_a \sqrt{\frac{\text{tr}((M_a^E)^2)}{d_E(d_E + 1)}} \\
 &\leq \frac{N(A)}{2\sqrt{d_E + 1}} \leq \frac{N(A)}{2\sqrt{d_E}}
 \end{aligned} \tag{B.3}$$

where we used  $\text{tr}((M_a^E)^2) \leq \|M_a\|^2 \text{tr}(\Pi^E) \leq d_E$ , and  $N(A)$  is the total number of outcomes, i.e. range of the sum over  $a$ . Inserting this bound in (B.1), we find

$$\langle D_{\mathcal{A}}(\psi, \Omega_E) \rangle_\psi \leq \sum_{A \in \mathcal{A}} \langle D_A(\psi, \Omega_E) \rangle_\psi \leq \frac{N(\mathcal{A})}{2\sqrt{d_E}} \tag{B.4}$$

where  $N(\mathcal{A})$  is the total number of different outcomes in the full set  $\mathcal{A}$ .

Inserting this bound in Levy's lemma allows to prove the first of the probability statements in theorem 3.2.

The second part of theorem 3.2 is proved by repeating the argument above for the time-averaged distinguishability as the function of  $\psi$  in Levy's lemma, and replacing  $M_a$  by  $e^{iHt} M_a e^{-iHt}$ :

$$\langle \langle D_{\mathcal{A}}(\psi(t), \Omega) \rangle_t \rangle_\psi = \lim_{T \rightarrow \infty} \frac{1}{T} \int dt \langle D_{\mathcal{A}}(\psi(t), \Omega) \rangle_\psi \leq \frac{N(\mathcal{A})}{2\sqrt{d_E}}. \tag{B.5}$$

This finishes the proof of of indistinguishability from equilibrium for almost all states at almost all times.

---

<sup>15</sup>If none of the observables in  $\mathcal{A}$  take states outside of  $\mathcal{H}_E$  there is no need to introduce  $M_a^E$ .

## C Proof of theorem 3.3

Let us expand the initial state of the apparatus  $|\alpha\rangle = \sum_j \alpha_j |j\rangle$  in the eigenbasis  $|j\rangle \in \mathcal{H}_A$  of the operator  $J$  appearing in the interaction Hamiltonian  $H_{\text{int}} = \lambda A \otimes J$ . The proof first consists in establishing a bound involving the real and imaginary components of  $S(t)$

$$S(t) = \langle \alpha | e^{-i\lambda(\delta a)Jt} | \alpha \rangle. \quad (\text{C.1})$$

Direct calculation allows us to find

$$\begin{aligned} \text{Re } S(t) &= \sum_j |\alpha_j|^2 \cos((\delta a)\lambda j t) \\ &\geq \sum_j |\alpha_j|^2 \left( 1 - \frac{2}{\pi} \left( (\delta a)\lambda j t + \sin(\lambda(\delta a)j t) \right) \right) \\ &= 1 - \frac{2(\delta a)\lambda \langle J \rangle t}{\pi} + \frac{1}{\pi} \text{Im } S(t), \end{aligned} \quad (\text{C.2})$$

where we used  $\cos(x) \geq 1 - \frac{2}{\pi}(x + \sin(x))$  in the second line. When we impose the orthonormality condition  $S(t) = 0$  (3.33), to guarantee the distinction between two different eigenvalues of the operator  $A$  separated by  $\delta a$ , we derive the inequality

$$\frac{\pi}{\lambda \langle J \rangle t} \leq \delta a, \quad (\text{C.3})$$

proving (3.34) holds.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] S.W. Hawking, *Particle creation by black holes*, *Commun. Math. Phys.* **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [[INSPIRE](#)].
- [2] S. Popescu, A.J. Short and A. Winter, *The foundations of statistical mechanics from entanglement: individual states vs. averages*, [[quant-ph/0511225](#)].
- [3] S. Popescu, A.J. Short and A. Winter, *Entanglement and the foundations of statistical mechanics*, *Nat. Phys.* **2** (2006) 754.
- [4] I. Heemskerk, J. Penedones, J. Polchinski and J. Sully, *Holography from conformal field theory*, *JHEP* **10** (2009) 079 [[arXiv:0907.0151](#)] [[INSPIRE](#)].
- [5] N. Arkani-Hamed, S. Dubovsky, A. Nicolis, E. Trincherini and G. Villadoro, *A measure of de Sitter entropy and eternal inflation*, *JHEP* **05** (2007) 055 [[arXiv:0704.1814](#)] [[INSPIRE](#)].
- [6] J. Simon, *Extremal black holes, holography and coarse graining*, *Int. J. Mod. Phys. A* **26** (2011) 1903 [[arXiv:1106.0116](#)] [[INSPIRE](#)].
- [7] J.M. Maldacena, *Eternal black holes in anti-de Sitter*, *JHEP* **04** (2003) 021 [[hep-th/0106112](#)] [[INSPIRE](#)].



- [8] V. Balasubramanian et al., *Typicality versus thermality: an analytic distinction*, *Gen. Rel. Grav.* **40** (2008) 1863 [[hep-th/0701122](#)] [[INSPIRE](#)].
- [9] V. Balasubramanian, J. de Boer, V. Jejjala and J. Simon, *The library of Babel: on the origin of gravitational thermodynamics*, *JHEP* **12** (2005) 006 [[hep-th/0508023](#)] [[INSPIRE](#)].
- [10] A.J. Short and T.C. Farrelly, *Quantum equilibration in finite time*, *New J. Phys.* **14** (2012) 013063 [[arXiv:1110.5759](#)].
- [11] P. Reimann, *Canonical thermalization*, *New J. Phys.* **12** (2010) 055027 [[arXiv:1005.5625](#)].
- [12] S. Lloyd, *Pure state quantum statistical mechanics and black holes*, [arXiv:1307.0378](#).
- [13] V.D. Milman and G. Schetchman, *Asymptotic theory of finite dimensional normed spaces*, *Lect. Notes Math.*, Vol. 1200, Springer-Verlag (1986).
- [14] M. Ledoux, *The concentration of measure phenomenon*, *AMS Mathematical Surveys and Monographs*, Vol. 89, American Mathematical Society (2001).
- [15] A.J. Short, *Equilibration of quantum systems and subsystems*, *New J. Phys.* **13** (2011) 053009 [[arXiv:1012.4622](#)].
- [16] N. Linden, S. Popescu, A.J. Short and A. Winter, *Quantum mechanical evolution towards thermal equilibrium*, *Phys. Rev. E* **79** (2009) 061103 [[arXiv:0812.2385](#)].
- [17] J. von Neumann, *Beweis des Ergodensatzes und des H-Theorems in der neuen Mechanik*, *Z. Phys.* **57** (1929) 30, English translation: *Proof of the ergodic theorem and the H-theorem in quantum mechanics*, *Eur. Phys. J. H* **35** (2010) 201 [[arXiv:1003.2133](#)].
- [18] S. Goldstein, J.L. Lebowitz, C. Mastrodonato, R. Tumulka and N. Zangh, *Normal typicality and von Neumann's quantum ergodic theorem*, *Proc. Roy. Soc. A* **466** (2010) 3203 [[arXiv:0907.0108](#)].
- [19] M.A. Nielsen and I.L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge U.K. (2010).
- [20] D.N. Page, *Average entropy of a subsystem*, *Phys. Rev. Lett.* **71** (1993) 1291 [[gr-qc/9305007](#)] [[INSPIRE](#)].
- [21] W.H. Zurek, *Decoherence, einselection and the quantum origins of the classical*, *Rev. Mod. Phys.* **75** (2003) 715 [[INSPIRE](#)].
- [22] J. von Neumann, *Mathematical foundations of quantum mechanics*, Vol. 2, Princeton University Press, Princeton U.S.A. (1996).
- [23] N. Margolus and L.B. Levitin, *The maximum speed of dynamical evolution*, *Physica D* **120** (1998) 188 [[quant-ph/9710043](#)] [[INSPIRE](#)].
- [24] V. Balasubramanian, D. Marolf and M. Rozali, *Information recovery from black holes*, *Gen. Rel. Grav.* **38** (2006) 1529 [*Int. J. Mod. Phys. D* **15** (2006) 2285] [[hep-th/0604045](#)] [[INSPIRE](#)].
- [25] J.M. Maldacena, *The large- $N$  limit of superconformal field theories and supergravity*, *Adv. Theor. Math. Phys.* **2** (1998) 231 [*Int. J. Theor. Phys.* **38** (1999) 1113] [[hep-th/9711200](#)] [[INSPIRE](#)].
- [26] O. Aharony, S.S. Gubser, J.M. Maldacena, H. Ooguri and Y. Oz, *Large- $N$  field theories, string theory and gravity*, *Phys. Rept.* **323** (2000) 183 [[hep-th/9905111](#)] [[INSPIRE](#)].
- [27] N. Lashkari, *Equilibration of small and large subsystems in field theories and matrix models*, [arXiv:1304.6416](#) [[INSPIRE](#)].

- [28] E. Witten, *Baryons in the  $1/N$  expansion*, *Nucl. Phys. B* **160** (1979) 57 [INSPIRE].
- [29] V. Balasubramanian, V. Jejjala and J. Simon, *The library of Babel*, *Int. J. Mod. Phys. D* **14** (2005) 2181 [hep-th/0505123] [INSPIRE].
- [30] K. Papadodimas and S. Raju, *The black hole interior in AdS/CFT and the information paradox*, *Phys. Rev. Lett.* **112** (2014) 051301 [arXiv:1310.6334] [INSPIRE].
- [31] K. Papadodimas and S. Raju, *State-dependent bulk-boundary maps and black hole complementarity*, *Phys. Rev. D* **89** (2014) 086010 [arXiv:1310.6335] [INSPIRE].
- [32] T. Banks, M.R. Douglas, G.T. Horowitz and E.J. Martinec, *AdS dynamics from conformal field theory*, hep-th/9808016 [INSPIRE].
- [33] I. Bena, *On the construction of local fields in the bulk of  $AdS_5$  and other spaces*, *Phys. Rev. D* **62** (2000) 066007 [hep-th/9905186] [INSPIRE].
- [34] A. Hamilton, D.N. Kabat, G. Lifschytz and D.A. Lowe, *Local bulk operators in AdS/CFT: a boundary view of horizons and locality*, *Phys. Rev. D* **73** (2006) 086003 [hep-th/0506118] [INSPIRE].
- [35] A. Hamilton, D.N. Kabat, G. Lifschytz and D.A. Lowe, *Holographic representation of local bulk operators*, *Phys. Rev. D* **74** (2006) 066009 [hep-th/0606141] [INSPIRE].
- [36] D. Kabat, G. Lifschytz and D.A. Lowe, *Constructing local bulk observables in interacting AdS/CFT*, *Phys. Rev. D* **83** (2011) 106009 [arXiv:1102.2910] [INSPIRE].
- [37] D. Kabat and G. Lifschytz, *CFT representation of interacting bulk gauge fields in AdS*, *Phys. Rev. D* **87** (2013) 086004 [arXiv:1212.3788] [INSPIRE].
- [38] K. Papadodimas and S. Raju, *An infalling observer in AdS/CFT*, *JHEP* **10** (2013) 212 [arXiv:1211.6767] [INSPIRE].
- [39] G. 't Hooft, *The black hole interpretation of string theory*, *Nucl. Phys. B* **335** (1990) 138 [INSPIRE].
- [40] L. Susskind, L. Thorlacius and J. Uglum, *The stretched horizon and black hole complementarity*, *Phys. Rev. D* **48** (1993) 3743 [hep-th/9306069] [INSPIRE].
- [41] E. Witten, *Anti-de Sitter space, thermal phase transition and confinement in gauge theories*, *Adv. Theor. Math. Phys.* **2** (1998) 505 [hep-th/9803131] [INSPIRE].
- [42] B. Czech, J.L. Karczmarek, F. Nogueira and M. Van Raamsdonk, *The gravity dual of a density matrix*, *Class. Quant. Grav.* **29** (2012) 155009 [arXiv:1204.1330] [INSPIRE].
- [43] V.E. Hubeny and M. Rangamani, *Causal holographic information*, *JHEP* **06** (2012) 114 [arXiv:1204.1698] [INSPIRE].
- [44] B. Czech, J.L. Karczmarek, F. Nogueira and M. Van Raamsdonk, *Rindler quantum gravity*, *Class. Quant. Grav.* **29** (2012) 235025 [arXiv:1206.1323] [INSPIRE].
- [45] M. Van Raamsdonk, *Evaporating firewalls*, arXiv:1307.1796 [INSPIRE].
- [46] J. Simon, *Small black holes versus horizonless solutions in AdS*, *Phys. Rev. D* **81** (2010) 024003 [arXiv:0910.3225] [INSPIRE].
- [47] S.D. Mathur, *Fuzzballs and black hole thermodynamics*, arXiv:1401.4097 [INSPIRE].
- [48] B. Collins and P. Śniady, *Integration with respect to the Haar measure on unitary, orthogonal and symplectic group*, *Commun. Math. Phys.* **264** (2006) 773 [math-ph/0402073].